Lowrie, Tom and Diezmann, Carmel (2010) *Solving graphics tasks : gender differences in middle-school students.* Learning and Instruction. (In Press)

Solving graphics tasks: Gender differences in middle-school students

Tom Lowrie [a], Carmel M. Diezmann [b]

[a] *School of Education, Charles Sturt University, Locked Bag 588, Wagga Wagga NSW 2678, Australia*
[b] *Centre for Learning Innovation, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia*

Abstract

The capacity to solve tasks that contain high concentrations of visual-spatial information, including graphs, maps and diagrams, is becoming increasingly important in educational contexts as well as everyday life. This research examined gender differences in the performance of students solving graphics tasks from the Graphical Languages in Mathematics (GLIM) instrument that included number lines, graphs, maps and diagrams. The participants were 317 Australian students (169 males and 148 females) aged 9 to 12 years. Boys outperformed girls on graphical languages that required the interpretation of information represented on an axis and graphical languages that required movement between two- and three-dimensional representations (generally Map language).

Keywords: Mathematics education; Gender differences; Graphics; Decoding

* Corresponding author. Tel.: +61 2 6933 4328; fax: +61 2 6933 2962
E-mail addresses: tlowrie@csu.edu.au (T. Lowrie); c.diezmann@qut.edu.au (C. M. Diezmann)

1. Introduction

The information age has provided new and increased demands on our capacity to represent, manipulate and decode information in diagrammatical and graphical forms (Lowrie & Diezmann, 2007). At an early age, students are required to make sense of graphical representations in a variety of contexts. It is certainly the case that the non-verbal processing of information, such as the interpretation of graphs, maps and drawings, is necessary in educational contexts as well as everyday life (Åberg-Bengtsson, 1999). Thus, attention to graphicacy is imperative (Åberg-Bengtsson & Ottosson, 2006), that is, «being 'graphicate' is becoming an important part of everyday knowledge, equal in status to being literate and numerate» (pp. 43-44).

Recent studies on the effect of graphics in problem solving have examined the extent to which learners use graphics and text to interpret information. This connectivity (or lack of) between text and graphics has been investigated in terms of cognitive load (Sweller, 1994), graphic design (Kosslyn, 2006) and whether or not the graphic contains essential information for solution (Elia, Gagatsis, & Demetriou, 2007; Gagatsis & Elia, 2004). Other researchers (Postigo & Pozo, 2004) have argued that previous research conducted in this field is quite heterogeneous since the study of maps, diagrams and numerical graphs have their own syntax and conventions. Nevertheless, some consistent findings have recently emerged. In a study that examined university-aged (M = 26 years) students' processing of meteorology diagrams (Schmidt-Weigand, Kohnert, & Glowalla, in press), participants spent more time examining text than interpreting dynamic visualisations. Interestingly, these adults predominantly focused on the text when interpreting the maps and tended to ignore the visual (and graphic) representations. Similarly, Berends and van Lieshout (2009) found that Grade 5 students' arithmetic performance decreased when they were required to find necessary information from graphics contained in word problems. In fact, students were more likely to achieve a correct solution, and finish the work in a timely manner, when they did not have to rely on essential graphical information. Students' lack of attention to graphics (as compared to text) in a mathematics task or their inability to utilize information from a graphic is apparent early in their schooling. For example, Gagatsis and Elia (2004) reported that only 18% of students in Grades 1 to 3 (N = 1447) were successful in using essential information from a number line on a one-step addition problem.

Gagatsis and Elia's (2004) study highlighted the dual role that graphics play in mathematics tasks. Graphics can be used as autonomous or auxiliary representations. An autonomous representation contains information essential to the task which is not presented elsewhere (i.e., in text or symbols). By contrast, an auxiliary representation contains information which might be helpful in problem solving (e.g., providing a cue to the context) but is not essential. Elsewhere, we have referred to the autonomous and auxiliary roles of graphics in terms of informational and contextual roles, respectively, and argued that students encounter graphics in both roles in mathematics instruction and assessment (Diezmann & Lowrie, 2008).

The present study examined the extent to which students are able to solve problems where the graphic plays an autonomous or informational role in the solution. To emphasise the essential role of the graphic in the solution, the selected items are rich in graphics but have limited text. Hence, for success on these items it is necessary for the solver to effectively interpret (and thus decode) the graphic and extract the essential information. Henceforth, we refer to these types of tasks as "graphics tasks". In addition, the study considers the role of gender on task performance given the key role gender appears to play on the decoding of graphics items (Lowrie & Diezmann, 2007).

1.1. Decoding graphics tasks

When decoding a graphic, individuals must contend with multiple sources of information which may include text, keys or legends, axes, and labels (Kosslyn, 2006), as well as perceptual elements of retinal variables (e.g., depth of shading and pattern) (Bertin, 1967/1983). It is therefore necessary to consider these components, which are often interrelated, in conjunction with the actual mathematics that is contained within a given task. Studies by Hittleman (1985) and Carpenter and Shah (1998) have shown that students find it challenging to move between text and graphics to the extent that it can disturb their thinking. Indeed, the graphic can often make the task more difficult to decode (Berends & van Lieshout, 2009; Elia et al., 2007; Schmidt-Weigand et al., in press). The degree of difficulty students experience with a graphic depends on the graphic itself. Baker, Corbett, and Koedinger (2001) demonstrated convincingly that, although graphics may be informationally equivalent, the particular graphic in use has a strong influence on students' success. Hence, if the student is not able to access and interpret the information effectively, the actual mathematics embedded within a given graphics task is not likely to be influential in the solution.

1.2. A graphical framework

The theoretical framework of the present study is derived from the work of Bertin (1967/1983), who describes graphics in terms of information within the graphic, the properties of the system, and the underlying components that govern and combine these properties. Of particular relevance to this study is the notion that graphic systems are made up of a number of variables that on their own or combined with other variables provide a classification for different types of graphics. Mackinlay (1999) built on Bertin's work to establish exemplars for graphics that are closely aligned to mathematics tasks. He argued that graphics can be categorised into six types of "graphical languages", which represent mathematical relationships among perceptual elements and use particular encoding techniques. These six graphical languages are named Axis, Apposed-position, Map, Retinal-list, Connection, and Miscellaneous, respectively.

<u>Axis</u> language is a single-position graphical language which encodes information by the position of a mark set on one axis. Such graphical language requires the decoding of information along either a horizontal or vertical axis—according to Mackinlay (1999), these graphical languages should be classified in the same manner despite their horizontal or vertical orientation.

<u>Apposed-position</u> language encodes information by a mark set that is positioned along both <u>x</u> and <u>y</u> axes. It is necessary to coordinate the information from both axes to generate the correct solution.

<u>Map</u> language, which has fixed positions, encodes information with graphical techniques that are specific to maps. This graphical language also requires the interpretation of information that is generally specific to map tasks. This information includes the interpretation of symbols in relation to position (e.g., position in relation to co-ordinates or position in relation to a bird's-eye perspective of orientation).

Connection language encodes information by connecting a set of node objects with a set of link objects. This graphical language requires the interpretation of information in relation to connections and links between sets of objects (e.g., family trees and tennis draws).

Retinal-list language does not require any form of position encoding since there is no requirement to interpret information contained along a continuum or within fixed point positions. Encoding is undertaken using visualisation and orientation processing—with tasks often requiring the rotation or reflection of objects.

Miscellaneous languages encode information with a variety of additional graphical techniques. These graphical languages include graphics such as pie charts and Venn diagrams. Decoding in Miscellaneous languages requires knowledge of the conventions of particular graphics. For example, interpreting a pie chart requires knowledge of how information is represented proportionally on a circle (for a discussion of students' strategies on a pie chart item see Diezmann & Lowrie, 2009b). However, even in tasks that are rich in graphics, the non-graphical information presented will also be part of the structure of the graphical languages used.

Essentially, the six graphical languages are described as a set of perceptual elements and encoding techniques. Apart from Miscellaneous languages, graphical languages have a unique graphical structure—Miscellaneous languages can have various graphical structures. Cleveland and McGill (1984) maintained that graphical information is more likely to be decoded in an accurate manner when information is presented along a common scale (as is the case with Axis, Apposed-position and some Map languages). Graphical languages are more difficult to decode when they are represented on a non-aligned scale; or use length, direction, angle, and shading (as in the case of Retinal-list, Connection and Miscellaneous languages). Thus, graphics within Axis and Apposed-position languages are generally easier to decode than those within the Retinal-list language. Elsewhere, we have argued that students need to appreciate how the visual components and spatial organisation of particular graphics impact on the interpretation of symbols within a graphic. For example, relative position is important in determining the numerical value of missing numbers of a number line (Diezmann & Lowrie, 2006).

1.3. Structure of graphical languages in standardised assessment

Graphical language structure involves all the representations that are presented to the individual when solving a graphics task. The structure of a graphical language includes not only the actual graphic, but all of the information embedded within the task (Kosslyn, 2006). In the present article it is accepted that the graphical language structure contains the graphic, any text, contextual information and other external representations. It requires the non-verbal processing of information. As a consequence, success on a graphics task can be attributed to one's capacity to navigate representations in multiple modes (Berends & van Lieshout, 2009). This viewpoint is advocated by others. Logan and Greenlees (2008) examined student performance on graphical languages that were almost identical in structure and reported that it was difficult to separate the graphicacy demand, which was embedded in a graphics task, from other demands (including mathematical content and linguistic demands).

A further influence on an individual's ability to interpret a graphic (e.g., a diagram) is the solver's prior knowledge which includes skills, preferences, and experiences (Brna, Cox, & Good, 2001). Moreover, it involves interaction between the encoding of information from the actual graphic and the more general processes of applying knowledge of conventions and domain knowledge (Canham & Hergarty, in press). Thus, understanding the performance of students on autonomous tasks rich in graphics requires an appreciation of the influences of other external representations and students' skills or prior knowledge.

1.4. Gender differences in mathematics

A broad body of literature has examined the performance differences between males and females on non-verbal or spatial tasks. Although performance differences are widely acknowledged (Linn & Petersen, 1985), the extent of these differences, the age when these differences occur (and/or diminish), and the nature of the tasks have raised considerable debate. Spelke (2005) in a comprehensive study indicated that the gap between the performance of boys and girls has diminished in the past ten years. Nevertheless, boys tend to perform better on tasks that require mental rotations or when tasks encourage the manipulation of objects in the mind (Spelke, 2005), and increasingly, such tasks are presented in mathematics tests (Diezmann & Lowrie, 2008; Lowrie & Diezmann, 2005). What has not been investigated is the specific type of tasks (in terms of the type of graphic) which produce these differences.

Many reasons for apparent performance differences between males and females on non-verbal and more generic mathematics tasks have emerged from the literature. Explanations include the confidence levels of girls (Forgasz, Leder, & Kaur, 2001); their attitudes toward mathematics (Forgasz, Leder, & Kloosterman, 2004); the fact that boys tend to process rotation tasks more quickly than girls (Wiedenbauer & Jansen-Osmann, 2008); and the notion that more boys than girls have extreme talent in mathematics (Benbow, 1988). Other accounts include the view that boys' and girls' everyday experiences are different (Tracey, 1990); the age of students (Levine, Huttenlocher, Taylor, & Langrock, 1999); and the manner in which tasks are represented (Lokan, Greenwood, & Creswell, 2001)—with short answer questions tending to advantage males.

Despite the abundance of studies related to gender, and due to the volume of generic studies being undertaken, Fennema and Leder (1990) have called on studies to be more focused and strategic when examining possible differences between the performance of males and females in mathematics. They suggested that rather than taking a broad view of mathematics performance, more studies should be framed at a micro level rather than across large populations. In a similar vein, Mills, Ablard, and Stumpf (1993) suggested that investigations should examine gender differences across subskills rather than studying differences in overall scores. It could be argued that these subskills should be contained within tasks with a high non-verbal requirement since it has long been hypothesised that non verbal or spatial reasoning is an important variable in the effect of gender differences in mathematics (Tartre, 1990). Similarly, Lokan et al. (2001) concluded that tasks which were saturated with diagrammatic information were more likely to be successfully solved by males—and consequently, tasks that demand high levels of non-verbal reasoning should be a focus of gender-related studies (Lowrie & Diezmann, 2007).

1.4.1. Other factors affecting gender differences in mathematics

Participant age is the strongest predictor regarding performance differences between males and females (Hyde, Fennema, & Lamon, 1990). In their meta-analysis of 100 studies, Hyde et al. (1990) found that males were more likely to outperform females on mathematics tasks in high school and beyond. As Bielinski and Davison (1998) indicated, the gender gap tends to favour females in elementary school, with no gender differences in the middle years, and with males outperforming females in high school.

Other studies have considered the influence of task complexity on performance. Penner's (2003) study of mathematics and science achievement tests revealed that boys outperform girls in difficult mathematical problem-solving items. Differences between males and females were not detected in easier items; however, differences increased when questions became more difficult. Interestingly, a study by Bielinski and Davison (1998) found that males tended to outperform females on the hardest items, while females tended to outperform males on the easiest items.

The studies mentioned above refer to gender differences in relation to task difficulty on mathematics tasks that typically measure general problem-solving skills. There is a paucity of research which examines performance differences of boys and girls in relation to task structure (Friel, Curcio, & Bright, 2001). In the present study graphical language structure was taken into consideration in addition to the cognitive demands required to solve these specific tasks.

1.5. The present study

The present study is part of a 3-year longitudinal study that was designed to enhance understanding of the development of primary students' ability to decode information graphics that represent mathematical information. The aim of the present study was to examine the decoding performance of students' solving graphics tasks over time with particular attention to the influence of gender on performance. The specific aims were (a) to establish whether there were gender differences in students' decoding performance in relation to the six graphical languages; (b) to determine whether there were gender differences in students' decoding performance over a 3-year period; and (c) to establish whether decoding performance (in relation to gender) is influenced by task difficulty (in terms of complexity).

To consider the influence graphics tasks have on students' decoding performance, the Mackinlay's (1999) model of graphical languages (i.e., categories) was used as a theoretical framework. This framework, initially derived from the work of Bertin (1967/1983), provides a perceptual basis for analysing students' decoding performance on graphics tasks. The present study expands upon the research literature by examining students in a specific field of mathematics education, that is, visual-spatial tasks that contain graphics.

1.5.1. Hypotheses

Two hypotheses were formulated for the present study. That there would be gender differences (in favour of boys) on graphics tasks that were classified as difficult (Hypothesis 1) and that these performance differences would be most evident across Map languages (Hypothesis 2).

2. Method

2.1. Participants

The participants (N = 317; female = 148, male = 169) were randomly selected from nine primary schools across two different states in rural and metropolitan areas of Australia. This sample excludes students who did not participate in the testing for the three consecutive years. The students mostly had English as their first language (94%), with less than 5% of students being classified as Indigenous. The economic status of the participants' families was typically middle class in areas with relatively low unemployment. The schools included six non-government and three government schools. The participants were investigated in the last three years of their primary education (age range 9-12 years). The participants were not involved in any treatment program throughout the study—they continued with the mandatory curriculum of their respective states.

2.2. Instruments

2.2.1. The GLIM test

The Graphical Languages in Mathematics (GLIM) test is a 36-item test (Diezmann & Lowrie, 2009a) developed to determine students' decoding performance for each of the six graphical languages (Mackinlay, 1999). Initially, a bank of 58 items that were typically administered to students in Grades 4, 5, and 6 was variously trialled with primary-aged children (N = 796) in order to select items that: (a) varied in complexity; (b) required substantial levels of graphical interpretation; and (c) conformed to reliability and validity measures. The items were selected from state, national and international year-level mathematics and science tests that had been administered to students in their final three years of primary school or to similarly aged students (e.g., Queensland School Curriculum Council, 2000a). A panel of expert mathematics educators (N = 5) independently categorised each item within the framework (reliability coefficient Cronbach's alpha 0.9) in relation to Mackinlay's graphical-language classification. Questions with high literacy demands or responses with high variance were removed from the pool of items. Further details about the construction of the GLIM instrument are described elsewhere (Diezmann & Lowrie, 2009a).

In its final form, the GLIM test comprised six items from each of the six language categories. Three subtests were produced that categorised graphical language items by difficulty. The easiest pair of items in each language formed the category labelled Easy Items (12 items; two per each graphical language), the second easiest pair of items formed the category labelled Moderate Items (12 items; two per each graphical language), and the most difficult pair of items from each language were categorised as the Difficult Items (12 items; two per each graphical language). Appendices A and B display the moderate and difficult items, respectively. This categorisation was developed from the first annual testing of the cohort. Table 1 provides an overview of the six graphical languages with example appendix numbers.

Insert Table 1 about here

Scoring of the decoding performance ranged from 0 (non-successful response) to 1 (successful response). Solution responses were set by the respective assessment agencies, and thus scoring reliability was accurate. A maximum score of 12 could be gained in each of the 3 difficulty categories (e.g., 2 items X one of each of the 6 graphical languages). Table 2 provides the means and standard deviations for the 12 items of each of the three categories of item difficulty—represented as decimals.

Insert Table 2 about here

The 36-item GLIM test was administered to the students approximately 12 months apart in Grades 4, 5 and 6. The participants completed the GLIM test in approximately 50 minutes within intact classes; each class comprised 24-31 participants.

2.2.2. The Raven's test

The Raven's Standard Progressive Matrices (SPM; Raven, Raven, & Court, 1998), which is a subset of Raven's Progressive Matrices, was administered to all the participants at the outset of the study in order to measure non-verbal ability and specifically ability to form perceptual relations. The test was administered to participants in accordance with the test's protocol. The measure was used as a covariate (performance of the Raven's Test score) since it controlled for non-verbal reasoning.

2.3. Analyses

(M)ANCOVAs were used for the statistical analyses. Tests for the homogeneity of variance, using the Lavene test for equality of variance, indicated that the assumption of homogeneity of variance was not violated since all interactions were nonsignificant ($p > .05$). With respect to internal consistency, measures for the reliability of the independent categories produced Cronbach's alpha coefficient between .81 and .83. Importantly, the correlation of the gender and non-verbal variables was weak, $r = .03$, $p > .05$. Finally, the homogeneity of regression slope assumption was not violated as the interaction between each independent category and its respective dependent variable was not statistically significant ($p > .05$).

3. Results

3.1. Gender differences in decoding performance in the six graphical languages

The first aim of the study was to establish whether there are gender differences in students' decoding performance on the GLIM test in relation to the six graphical languages. In this analysis, the participants' correct responses were calculated over a 3-year period (Grades 4-6). Thus, means of the scores of the GLIM test were generated for a combined total for each of the six graphical languages (with possible totals ranging from 0-18 for each language). A repeated measures multivariate analysis of covariance with gender as between subjects factor, the six graphical languages as within subjects factor and score on Raven's test as covariate revealed statistically significant differences between the performance of boys and girls across the six graphical languages, $F(6, 309) = 9.37$, $p < .001$, partial $\eta^2 = .15$, with a significant effect for the covariate, $F(6, 309) = 50.47$, $p < .001$, partial $\eta^2 = .50$. Subsequent post hoc analysis revealed statistically significant differences across the gender variable for the Axis language, $F(1, 326) = 13.1$, $p < .001$, partial $\eta^2 = .11$, and the Map language, $F(1, 326) = 3.99$, $p < .001$, partial $\eta^2 = .03$. The gender effect for the other four categories was not significant. Across each of the six graphical languages, the boys' mean scores were higher than that of the girls (see Table 3) with this trend most evident across the Axis and Map languages.

Insert Table 3 about here

3.2. Gender differences in decoding performance over time

The second aim of the study was to determine whether there were gender differences in students' decoding performance on the GLIM test over a 3-year period (Grades 4-6). The means and standard deviations for the participants' scores of the GLIM test across the six graphical languages are presented in Table 4 (with possible totals ranging from 0-18 for each language) and indicate significant improvements in the students' performance on these graphical languages over time. The mean scores for the male students were higher than that of the female students in all six graphical languages in Grade 4, Grade 5 and Grade 6.

Insert Table 4 about here

To determine whether there were statistically significant differences between the decoding performances of males and females across the six graphical languages over time, a 3(year) x 2(gender) x 6(graphical languages) MANCOVA with Raven's test score as covariate was conducted. The analysis revealed a statistically significant covariate, Pillai's trace = .404, $\underline{F}(6, 1020) = 115.38$, $\underline{p} < .001$, partial $\eta^2 = .40$. There were statistically significant main effects for both the year, Pillai's trace = .237, $\underline{F}(6, 1020) = 22.92$, $\underline{p} < .001$, partial $\eta^2 = .12$, and gender, Pillai's trace = .682, $\underline{F}(6, 1020) = 15.23$, $\underline{p} < .001$, partial $\eta^2 = .40$. The interaction of gender with year was not significant, Pillai's trace = .016, $\underline{F}(6, 2040) = 1.34$, $\underline{p} = .19$, <u>ns</u>. With respect to year, subsequent post-hoc analyses revealed significant differences between students' performance across all six graphical languages (see Table 5 for $\underline{F}$ values and $\underline{p}$ levels). In relation to gender, there were statistically significant differences between boys and girls for the Axis language, $\underline{F}(1, 1032) = 52.3$, $\underline{p} < .001$, partial $\eta^2 = .07$, and the Map language, $\underline{F}(1, 1032) = 16.48$, $p < .001$, partial $\eta^2 = .02$.

3.3. Gender differences on decoding performance by category of item difficulty

The third aim of the study was to investigate whether decoding performance on the GLIM test (in relation to gender) is influenced by item difficulty (in terms of complexity). In this analysis three difficulty variables were generated. Specifically, each difficulty variable contained the respective two items from each of the six graphical languages. The maximum score for each variable was 12. The means and standard deviations for the boys and girls across the three years of the study (by category of item difficulty) are displayed in Table 5.

Insert Table 5 about here

The descriptive statistics show that males outperformed females in each of the three years across the three categories of item difficulty. The 3(categories) x 3(years) x (2gender) MANCOVA with Raven's test score as covariate showed a significant main effect of category, Pillai's trace = .406, $\underline{F}(3, 1023) = 233.5$, $\underline{p} < .001$, partial $\eta^2 = .40$. The MANCOVA revealed statistically significant main effect for year, Pillai's trace = .226, $\underline{F}(6, 2048) = 43.48$, $\underline{p} < .001$, partial $\eta^2 = .03$, and for gender, Pillai's trace = .026, $\underline{F}(3, 1023) = 9.13$, $\underline{p} < .001$, partial $\eta^2 .= .11$. The interaction of year with gender was not significant, Pillai's trace = .004, $\underline{F}(6, 2048) = .74$, $\underline{p} = .29$, <u>ns</u>. Post hoc analysis revealed statistically significant gender differences on difficult items, $F(2, 1032) = 18.99$, $p < .001$, partial $\eta^2 = .10$, and on moderate items, $F(2, 1032) = 14.77$, $p < .001$, partial $\eta^2 = .19$, but not on the easy items, $F(2, 1032) = .06$, $p > .05$.

Further analysis was undertaken to determine gender differences by item for difficult and moderate categories. Post hoc analysis revealed statistically significant differences between boys and girls on five items in the difficult category and four items in the moderate category, with boys outperforming girls in each instance (alpha levels were adjusted to $p = .004$ using the Bonferroni correction method). Table 6 provides means, standard deviations and F values for the nine statistically significant items in the moderate and difficult categories.

Insert Table 6 about here

Most of these items were either Axis (3 items) or Map (3 items) language items with two of them in each language from the difficult category. The other three items were Apposed-position, Retinal-list, and Connection languages items, respectively, two from the moderate category and one from the difficult category respectively.

4. Discussion

In the present study the focus was on student performance on graphics tasks that required specific types of graphical decoding. Interestingly, student performance increased (across all six graphical languages) in each year of the study. Such consistent increases in performance are important to highlight since older students do not always outperform younger students on graphic tasks with a "plateau effect" occurring at around Grade 5 (Diezmann, 2005). These results support the findings of other studies which showed that adolescents' (aged 12-16 years) graphing performance (Postigo & Pozo, 2004) and primary-aged children's' (aged 5-12 years) mapping skills (Liben & Downs, 1993) improved over time.

4.1. Gender differences on Axis language items

The present study showed decoding-performance differences in favour of boys on three of the four moderate and difficult items and are consistent with the results of studies by Hannula (2003) and Lowrie and Diezmann (2005), who found gender differences on Axis tasks in favour of boys for fifth-grade and fourth-grade students respectively.

The Axis language items required the decoding of information along either a horizontal or vertical continuum. The items with a horizontal orientation contained information along a horizontal continuum with necessary data contained within segments (rather than simply considering information from a start to end point). Item 2 required the problem solver to proportionalise the segments on the number line, thus creating the appropriate scale before making the distance calculation. Although Item 5 is represented vertically, and required the interpretation of a scale that descended from zero, the gender differences were still evident. In these items, elements of the graphics did not necessarily assist the problem solver (e.g., pictures of the insects or fruit). In fact, these additional elements of the graphic may have been distracting, as was the case with Berends and van Lieshout's (2009) study. Bertin (1967/1983) described such graphical representations in relation to an efficiency criterion, where the representations that surround a question can be processed or constructed in either an efficient or inefficient manner. He argued that the differences in perception time—the time taken to process information, such as to decipher how important the images are in relation to the question—can affect performance (Bertin, 1967/1983, p. 9). In other words, the graphics may actually act as distracters that can inhibit effective performance (Antonietti, 1991; Schmidt-Weigand et al., in press). These Axis language items required the students to move simultaneously between concrete and dynamic imagery (Presmeg, 1986) as they decoded or encoded information. The only Axis language item that did not produce a gender difference had a common "number line" representation that did not include any other symbols or words contained within the graphic (e.g., names or pictures) (Item 3). This finding is consistent with the outcomes of a study conducted by Liben and Downs (1993), which revealed no gender differences when Grade 5 children were required to complete a series of Axis language questions. It seems that the addition of other graphical information increased the demand on the performance of girls in relation to that of boys. That is, the additional text and contextual information associated with the graphic made the task more complex for girls.

4.2. Gender differences on Map language items

With respect to performance on Map items, Kitchin (1996) postulated that gender differences in the interpretation and decoding of maps may be a result of the females having less access to situations that develop spatial skills or that measuring tasks favour male problem-solving strategies. Boardman (1990) highlighted the fact that gender difference in mapping ability may increase over time and that by adolescence boys demonstrate more highly developed map skills than girls. In the present study, performance differences between boys and girls remained relatively constant over the 3-year period. It could be argued that the girls in our study were much more likely to be exposed to maps than students in the earlier studies—since these studies are more than a decade old given the abundance of maps in society. Today, there is increased attention given to maps in the school curriculum and arguably even more influential is the exposure all students have to maps in everyday life. For example, maps are increasingly displayed through the media (e.g., weather forecasts), on signs and visual displays, hand-held games (e.g., Gameboys) and even car navigation systems. Despite this increased exposure, gender differences remain.

The Map language items which revealed distinct differences between the performance of boys and girls all required the evoking of concrete and dynamic imagery—in these cases the capacity to interpret information from a bird's-eye view perspective. Item 16 required the manipulation of objects in the mind's eye (Kosslyn, 1983), with the rotation of three-dimensional (3D) and two-dimensional (2D) objects. The other two items required the interpretation of maps using directional processing and following directions using Euclidian (e.g., North, right) (Item 17) and non-Euclidian terminology (e.g., from the gate to the tap) (Item 15). Interestingly, girls tend to rely on landmark navigation more than boys (Spelke, 2005), which may have been a limiting factor in their capacity to decode these tasks. As Silverman and Choi (2006) found, females tend to use more holistic typographical approaches to solve graphics tasks. Item 14 was the only Map language item that did not produce a significant gender difference. Noteworthy is the relative ease of this item ($\underline{M}$ = .74) in comparison to other Map language items in this analysis.

4.3. Gender differences across other graphical languages items

Three additional items revealed gender differences in favour of boys. Item 10 (a Retinal-list language item) required the problem solver to consider a 2D (bird's-eye view) perspective of a 3D object. The cognitive processing required to solve this item was similar to the demands placed on the students to solve one of the Map language items (Item 16). Both items required movement between 2D and 3D processing. It is important to note that there were no gender differences on items that required 2D to 2D processing. There were gender differences in one of the Connection language items (Item 20). This particular item was the only connection item that required the linking of objects and nodes in a horizontal manner (with information being processed in either a left-to-right or right-to-left manner). This type of processing is akin to the processing required in the interpretation of Axis language items—where the gender differences were evident. The other item which revealed gender differences was an Apposed-position language item (Item 8). This item required similar processing to other Apposed-position language items, and as a result, the performance differences are difficult to explain. We suggest that this particular item required the problem solver to go beyond the surface features of the graphic and make inferences from the data (Cuoco, 2001) in ways not required with any other item within this language. Thus, the solution could not be derived from the graphic without additional processing of information. This complexity level may account for performance differences across gender.

4.4. Limitations of the study

The present study would be strengthened with the addition of more items which have similar structure and nature to the GLIM items that revealed performance differences between boys and girls. A study that presented students with additional Axis language items (with several tasks of both horizontal and vertical orientation) and additional Map language items (especially those that required students to decode directional tasks) would allow for a more detailed analysis of performance differences.

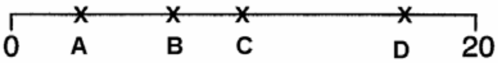
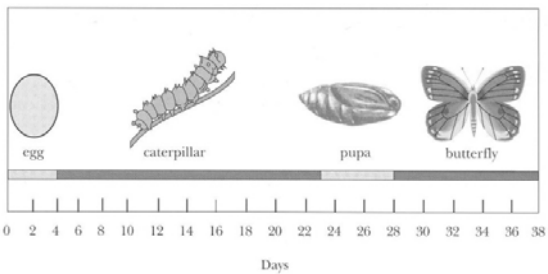
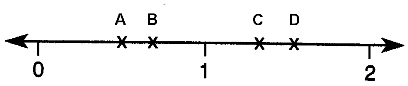
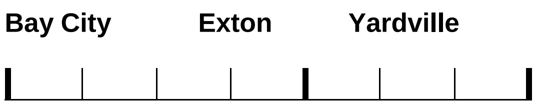
From a theoretical perspective, further studies that focus on gender differences should attempt to control for additional learning dimensions in order to better attribute psychological or cognitive differences among students. In the present study we controlled for general non-verbal ability; however in order to gain insights into explanations for these dramatic performance differences other cognitive capacities, in particular, need to be included in the design. Measures of cognitive capacity (particularly general mathematics ability) and spatial reasoning (specifically mental rotation and visual imagery) should be included as covariates in new studies.

4.5. Concluding comments

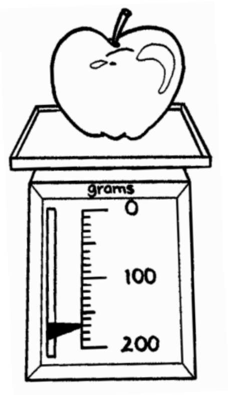
One of the central concerns of the study was to determine whether there would be gender differences on graphics tasks that required more complex levels of graphical decoding (Hypothesis 1). The results showed that boys outperformed girls on the more difficult tasks. Thus, hypothesis 1 is verified. These findings support previous research in relation to complex tasks (including Bielinski & Davison, 1998; Penner, 2003); however, the fact that girls did not outperform boys on easier tasks contradicts the findings of Bielinski and Davison (1998). We propose that as the complexity of the task structure increased (that is, the connectivity between graphic, text and contextual information), the more efficient boys were at navigating and decoding these representational forms.

Our second hypothesis was framed around the view that boys would outperform girls on Map language items (Hypothesis 2). There were, in fact, statistically significant differences between the performance of boys and girls (in favour of boys) across the Map language in each of the three years of the study. We suggest that the performance differences are associated with item structure and specifically graphical representations that required vertical and/or horizontal decoding of information. In relation to hypothesis 2, it is noteworthy that were even greater differences between the performance of boys and girls on Axis items than that of Map items. We conclude that many of the Axis and Map language items encouraged processing that was directional (e.g., North, South, right, left), or along a single-axis continuum with information processed vertically or horizontally.

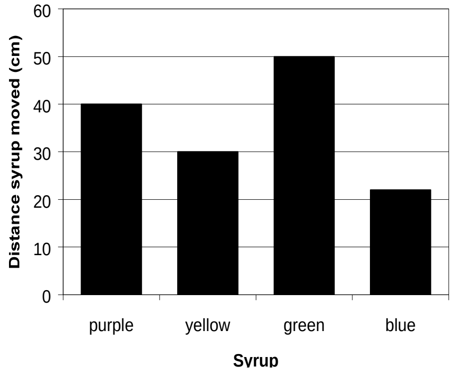
Appendix

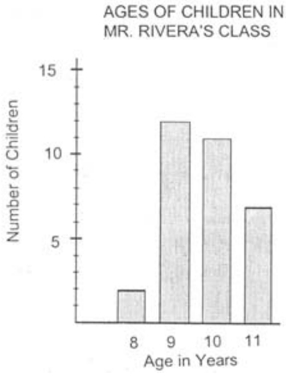
Estimate where you think 17 should go on this number line.  **Item 1 – Axis, Easy** Queensland School Curriculum Council. (2000a). *Aspects of numeracy test:Year 3*, p. 11.	The following graph shows the length of time taken for the four stages in the life of a butterfly.  How many days are there in the caterpillar stage? **Item 2 – Axis, Moderate** Educational Testing Centre. (2001a). *Australian schools science competition: Year 5*, p. 2.
Estimate where you think 1.3 should go on this number line.  **Item 3 – Axis, Moderate** Queensland School Curriculum Council. (2000b). *Aspects of numeracy test: Year 7*, p. 8.	 On the road shown above, the distance from Bay City to Exton is 60 kilometres. What is the distance from Bay City to Yardville? ☐ 45 kilometres ☐ 75 kilometres ☐ 90 kilometres ☐ 105 kilometres **Item 4 – Axis, Difficult** National Center for Education Statistics, US Department of Education. (n.d.). *2003 NAEP questions: Year 4*, p. 19.

<table>
<tr>
<td>
What is the mass of the apple?



Item 5 – Axis, Difficult
Queensland School Curriculum Council. (2001a). *Aspects of numeracy test: Year 3*, p. 14.
</td>
<td>
Syrups are thick, sticky liquids. The thicker the syrup, the slower it will move down a slope.
The graph shows the distance four different syrups moved down a slope in one minute.
Which syrup is the thickest?

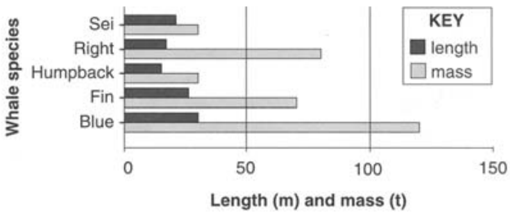


Item 6 – Apposed-position, Easy
Educational Testing Centre. (2003). *Australian schools science competition: Year 4*, p. 3.
</td>
</tr>
<tr>
<td>


The graph above shows how many of the 32 children in Mr Rivera's class are 8, 9, 10 and 11 years old. Which of the following is true?

Most are younger than 9, Most are younger than 10, Most are 9 or older, None of the above is true.

Item 7 – Apposed-position, Moderate
National Center for Education Statistics, US Department of Education. (n.d.). *2003 NAEP questions: Year 4*, q. 229.
</td>
<td>
The graph compares the maximum length and mass to which some whales grow.

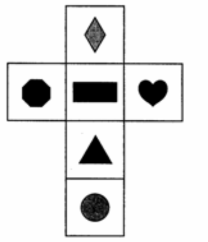


A fisherman reported that a whale 25 metres long and weighing approximately 80 tonnes had beached itself.

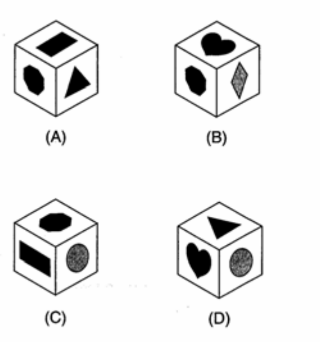
Which species of whale could this be?

Item 8 – Apposed-position, Difficult
Educational Testing Centre. (2002b). *Australian schools science competition: Year 6*, p. 6.
</td>
</tr>
</table>

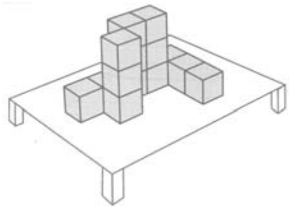
This is the net of a cube.



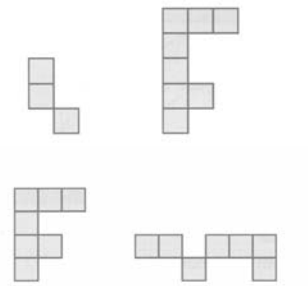
Which one of these cubes could be made by folding the net?



Item 9 – Retinal-list, Easy
Educational Testing Centre. (2002c). *Primary school mathematics competition: Year 4*, p. 9.

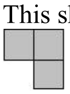


What does this model look like from above?

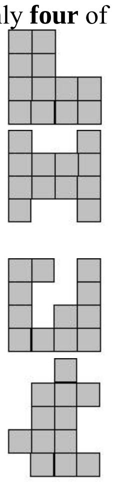


Item 10 – Retinal-list, Moderate
Queensland Studies Authority. (2002b). *Aspects of numeracy test: Year 5*, p. 10.

This shape was used to make different designs.

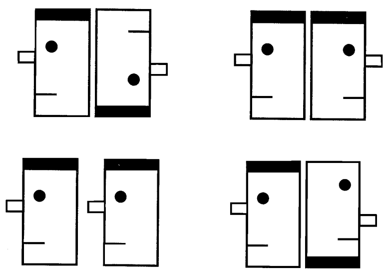


Which of the following designs **cannot** be made using only **four** of the shapes above?



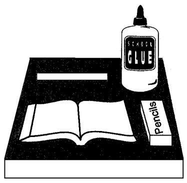
Item 11 – Retinal-list, Moderate
Educational Testing Centre. (2001c). *Primary school mathematics competition: Year 4*, p. 8.

Which two faces show a flip?

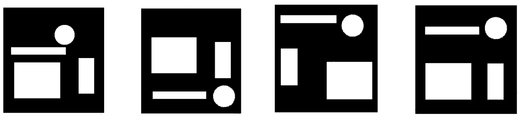


Item 12 – Retinal-list, Difficult
Educational Testing Centre. (1995). *Australian schools science competition: Year 7*, p. 4.

Jasmine has a book, ruler, pencil case and glue on her desk.

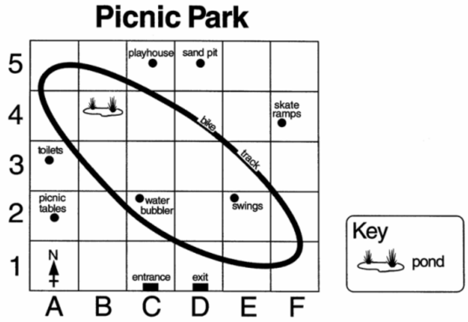


Which map best shows where everything is on Jasmine's desk?



(A) (B) (C) (D)

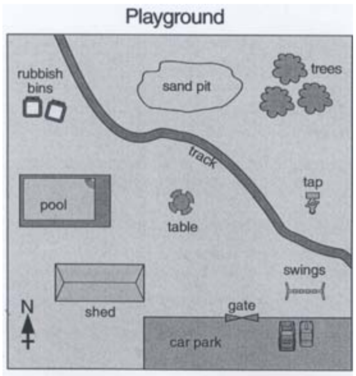
Item 13 – Map, Easy
Educational Testing Centre. (2002d). *Primary school mathematics competition: Year 5*, p. 4.



Deb rides her bike along the bike track. What part of the Park won't she ride through?

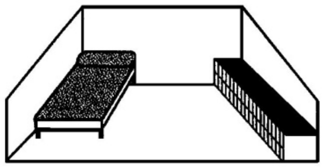
Item 14 – Map, Moderate
Queensland School Curriculum Council. (2001b). *Aspects of numeracy test: Year 5*, p. 16.

Ben went from the gate to the tap, then to the shed, then to the rubbish bins.
How many times did he cross the track?

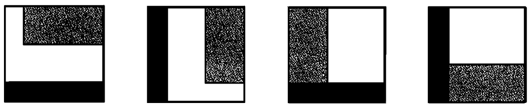


Item 15 – Map, Moderate
Queensland Studies Authority. (2002a). *Aspects of numeracy test: Year 3*, pp. 11, 3 (of insert).

Here is a bedroom.



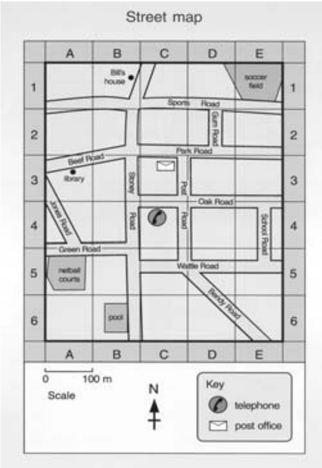
Which map shows the bedroom?



A B C D

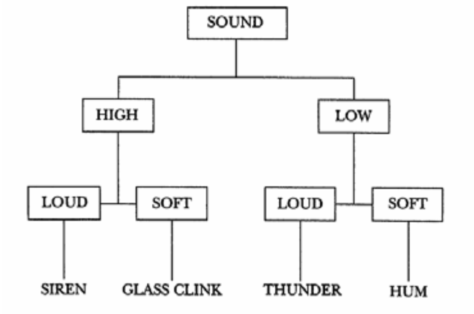
Item 16 – Map, Difficult
Queensland Studies Authority. (2002a). *Aspects of numeracy test: Year 3*, pp. 9, 1 (of insert).

Bill leaves the pool. He drives north and takes the first road on the right, then the second road on the left.
Which road is he in?



Item 17 – Map, Difficult
Queensland Studies Authority. (2002b). *Aspects of numeracy test: Year 5*, pp. 7, 3 (of insert).

This flowchart shows a way to describe sounds.

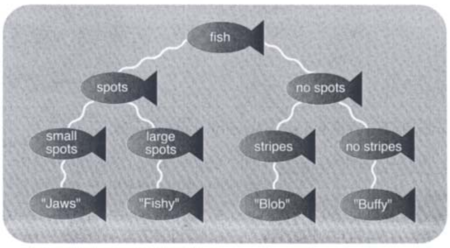


Which of the following describes a hum?

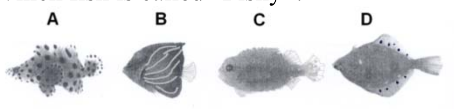
- ▢ high and loud
- ▢ high and soft
- ▢ low and loud
- ▢ low and soft

Item 18 – Connection, Easy
Educational Testing Centre. (2001b). *Australian schools science competition: Year 6*, p. 3.

David uses this key to tell his fish apart.

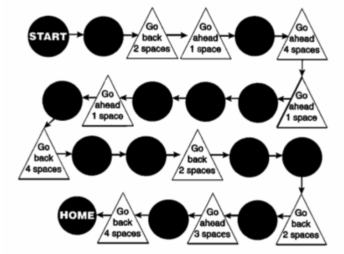


Which fish is called "Fishy"?



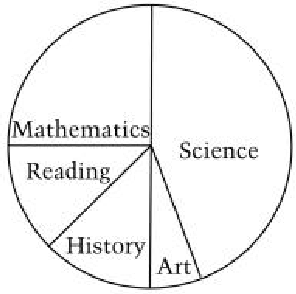
Item 19 – Connection, Moderate
Educational Testing Centre. (2002a). *Australian schools science competition: Year 3*, p. 4.

Two children are playing a board Game. They toss a standard dice and move forward the number of spaces to match the number on the dice.



What is the least number of tosses of the dice needed to reach ?

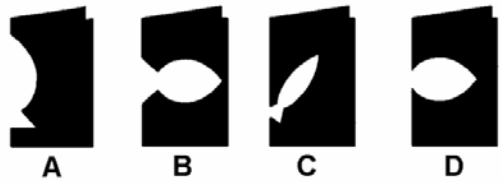
Item 20 – Connection, Difficult
Educational Testing Centre. (2002c). *Primary school mathematics competition: Year 4*, p. 9.



The pie chart above shows the portion of time Pat spent on homework in each subject last week. If Pat spent 2 hours on mathematics, about how many hours did Pat spend on homework altogether?

Item 21 – Miscellaneous, Easy
National Center for Education Statistics, US Department of Education. (n.d.). *2003 NAEP questions: Year 4*, q. 3.

Which piece of paper will show only this shape  when it is unfolded?



Item 22 – Miscellaneous, Moderate
Educational Testing Centre. (2001d). *Primary school mathematics competition: Year 5*, p. 2.

Which date is 3 weeks before 29 May?

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Item 23 – Miscellaneous, Difficult
Queensland Studies Authority. (2002b). *Aspects of numeracy test: Year 5*, pp. 5, 1 (of insert).

Acknowledgments

We would like to thank Tracy Logan and Kim Woodland for their insights, support and advice on the collaboration of data and dissemination of research for this article.

The work represented in this paper was supported by an Australian Research Council (ARC) Grant [DP0453366]. The views and ideas drawn from the research are that of the authors and not the ARC.

References

Åberg-Bengtsson, L. (1999). Dimensions of performance in the interpretation of diagrams, tables and maps: Some gender differences in the Swedish scholastic aptitude test. Journal of Research in Science Teaching, 36, 565-582.

Åberg-Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. Journal of Research in Science Teaching, 43, 43-62.

Antonietti, A. (1991). Why does mental visualization facilitate problem solving? In R. H. Logie & M. Denis (Eds.), Mental images in human cognition (pp. 211-227). Amsterdam, The Netherlands: Elsevier.

Baker, R. S., Corbett, A. T., Koedinger, K. R. (2001). Toward a model of learning data representations. In J. D. Moore & K. Stenning (Eds.), Proceedings of the 23rd annual conference of the Cognitive Science Society (pp. 45-50). Mahwah, NJ: Lawrence Erlbaum Associates.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. Behavioral and Brain Sciences, 11, 169-232.

Berends, I. E., & van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. Learning and Instruction, 19, 345-353.

Bertin, J. (1983). Semiology of graphics (W. J. Berg, Trans.). Madison, WI: The University of Wisconsin Press. (Original work published 1967)

Bielinski, J., & Davison, M. L. (1998). Gender differences by item difficulty: Interactions in multiple-choice mathematics items. American Educational Research Journal, 35, 455-476.

Boardman, D. (1990). Graphicacy revisited: Mapping abilities and gender differences. Educational Review, 42, 57-64.

Brna, P., Cox, R., & Good J. (2001). Learning to think and communicate with diagrams: 14 questions to consider. Artificial Intelligence Review, 15, 115-134.

Canham, M., Hergarty, M. (in press). Effects of knowledge and display design on comprehension of complex graphics. Learning and Instruction, doi:10.1016/j.learninstruc.2009.02.014

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. Journal of Experimental Psychology: Applied, 4, 75-100.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association, 79(387), 531-554.

Cuoco, A. A. (2001). The role of representation in school mathematics: National Council of Teachers of Mathematics yearbook. Reston, VA: National Council of Teachers of Mathematics.

Diezmann, C. M. (2005). Assessing primary students' knowledge of networks, hierarchies and matrices using scenario-based tasks. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonagh, R. Pierce, & A. Roche (Eds.), Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia (Vol. 1, pp. 289-296). Sydney, Australia: Mathematics Education Research Group of Australasia.

Diezmann, C. M., & Lowrie, T. (2006). Primary students' knowledge and errors on number lines. In P. Grootenboer, R. Zevenbergen, & M. Chinnappan (Eds.), Proceedings of the 29th annual conference of the Mathematics Education Research Group of Australasia (Vol. 1, pp. 171-178). Sydney, Australia: Mathematics Education Research Group of Australasia.

Diezmann, C. M., & Lowrie, T. (2008). The role of information graphics in mathematical proficiency. In M. Goos, R. Brown, & K. Makar (Eds.), Navigating currents and charting directions – Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia (Vol. 2, pp. 647-650). Brisbane, Australia: Mathematics Education Research Group of Australasia.

Diezmann, C. M., & Lowrie, T. (2009a). An instrument for assessing primary students' knowledge of information graphics in mathematics. Assessment in Education: Principles, Policy and Practice, 16(2), 131-147.

Diezmann, C. M., & Lowrie, T. (2009b). The role of fluency in a mathematics item with an embedded graphic: Interpreting a pie chart. ZDM: The International Journal on Mathematics Education, 41(5), 651-662.

Educational Testing Centre. (1995). Australian schools science competition: Year 7. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2001a). Australian schools science competition: Year 5. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2001b). Australian schools science competition: Year 6. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2001c). Primary school mathematics competition: Year 4. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2001d). Primary schools mathematics competition: Year 5. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2002a). Australian schools science competition: Year 3. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2002b). Australian schools science competition: Year 6. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2002c). Primary schools mathematics competition: Year 4. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2002d). Primary school mathematics competition: Year 5. Sydney, Australia: University of New South Wales.

Educational Testing Centre. (2003). Australian schools science competition: Year 4. Sydney, Austalia: University of New South Wales.

Elia, I., Gagatsis, A., & Demetriou, A. (2007). The effects of different modes of representation on the solution of one-step additive problems. Learning and Instruction, 17, 658-672.

Fennema, E., & Leder, G. S. (1990). Mathematics and gender. New York: Teachers College Press.

Forgasz, H. J., Leder, G. C., & Kaur, B. (2001). Who can(t) do maths—boys/girls? Beliefs of Australian and Singaporean secondary school students. Asia Pacific Journal of Education, 21(2), 106-116.

Forgasz, H. J., Leder, G. C., & Kloosterman, P. (2004). New perspectives on the gender stereotyping of mathematics. Mathematical Thinking and Learning, 6, 389-420.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. Journal for Research in Mathematics Education, 32(2), 124-158.

Gagatsis, A., & Elia, I. (2004). The effects of different modes of representation on mathematical problem solving. In M. Johnsen Høines & A. B.

Fuglestad (Eds.), Proceedings of the 28th conference of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 447-454). Bergen, Norway: Psychology of Mathematics Education.

Hannula, M. S. (2003). Locating fractions on a number line. In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), Proceedings of the 27th conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th conference of PME-NA (Vol. 3, pp. 3-24). Honolulu, HI: Psychology of Mathematics Education.

Hittleman, D. R. (1985). A picture is worth a thousand words…if you know the words. Childhood Education, 61-62, 32-36.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin, 107, 139-155.

Kitchin, R. (1996). Are there sex differences in geographic knowledge and understanding? The Geographical Journal, 162, 273-286.

Kosslyn, S. M. (1983). Ghosts in the mind's machine. New York: Norton.

Kosslyn, S. M. (2006). Graph design for the eye and mind. New York: Oxford University Press.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. Developmental Psychology, 35, 940-949.

Liben, L., & Downs, R. (1993). Understanding person-space-map relations: Cartographic and developmental perspectives. Developmental Psychology, 29, 739-752.

Linn, M., & Petersen, A. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development, 56, 1479-1498.

Logan, T., & Greenlees, J. (2008). Standardised assessment in mathematics: The tale of two items. In M. Goos, R. Brown, & K. Makar (Eds.), Navigating currents and charting directions – Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia (Vol. 2, pp. 655-658). Brisbane, Australia: Mathematics Education Research Group of Australasia.

Lokan, J., Greenwood, L., & Cresswell, J. (2001). 15-up and counting, reading, writing, reasoning…how literate are Australia's students? The PISA 2000 survey of students reading, mathematical and scientific literacy skills. Camberwell, Australia: Australian Council for Educational Research.

Lowrie, T., & Diezmann, C. M. (2005). Fourth-grade students' performance on graphical languages in mathematics. In H. L. Chick & J. L. Vincent (Eds.), Proceedings of the 30th annual conference of the International Group for the Psychology of Mathematics Education (Vol 3, pp. 265-272). Melbourne, Australia: Psychology of Mathematics Education.

Lowrie, T., & Diezmann, C. M. (2007). Solving graphics problems: Student performance in the junior grades. The Journal of Educational Research, 100(6), 369-377.

Mackinlay, J. (1999). Automating the design of graphical presentations of relational information. In S. K. Card, J. D. Mackinlay, & B. Schneiderman (Eds.), Readings in information visualization: Using vision to think (pp. 66-81). San Francisco, CA: Morgan Kaufmann.

Mills, C., Ablard, K., & Stumpf, H. (1993). Gender differences in academically talented young students' mathematical reasoning: Patterns across age and subskills. Journal of Educational Psychology, 85, 340-346.

National Center for Education Statistics (n.d.). 2003 NAEP questions: Year 4. Washington, DC: Institute of Education Sciences: The United States Department of

Education. Retrieved January 10, 2003, from http://nces.ed.gov/nationsreportcard/ITMRLS/search.asp?picksubj=Mathematics

Penner, A. (2003). International gender x item difficulty interactions in mathematics and science achievement tests. Journal of Educational Psychology, 95, 650-655.

Postigo, Y., & Pozo, J. (2004). On the road to graphicacy: The learning of graphical representation systems. Educational Psychology, 24, 623-644.

Presmeg, N. C. (1986). Visualization in high school mathematics. For the Learning of Mathematics, 6(3), 42-46.

Queensland School Curriculum Council. (2000a). 2000 Queensland Year 3 test: Aspects of numeracy. Camberwell, Australia: Australian Council for Educational Research.

Queensland School Curriculum Council. (2000b). 2000 Queensland Year 7 test: Aspects of numeracy. Camberwell, Australia: Australian Council for Educational Research.

Queensland School Curriculum Council. (2001a). 2001 Queensland Year 3 test: Aspects of numeracy. Camberwell, Australia: Australian Council for Educational Research.

Queensland School Curriculum Council. (2001b). Queensland Year 5 test: Aspects of numeracy. Camberwell, Australia: Australian Council for Educational Research.

Queensland Studies Authority. (2002a). 2002 Queensland Year 3 test: Aspects of numeracy. Camberwell, Australia. Australian Council for Educational Research.

Queensland Studies Authority. (2002b). 2002 Queensland Year 5 test: Aspects of numeracy. Camberwell, Australia: Australian Council for Educational Research.

Raven, J., Raven, J. C., & Court, J. H. (1998). Manual for Raven's standard progressive matrices. Oxford, England: Oxford Psychologists Press.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (in press). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. Learning and Instruction, doi: 10.1016/j.learninstruc.2009.02.11

Silverman, I., & Choi, J. (2006). Non-Euclidean navigational strategies of women: Compensatory response or evolved dimorphism? Evolutionary Psychology, 4, 75-84.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. American Psychologist, 60, 950-958.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4, 295-312.

Tartre, L. A. (1990). Spatial skills, gender, and mathematics. In E. Fennema & G. S. Leder (Eds.), Mathematics and gender (pp. 27-59). New York: Teachers College Press.

Tracey, D. A. (1990). Toys, spatial ability, and science and mathematics achievement: Are they related? Sex Roles, 17, 115-137.

Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. Learning and Instruction, 18, 30-41.

Table 1
An overview of the six graphical languages with example appendix number

Graphical languages	Examples	Encoding technique	Appendix number of easy, moderate and difficult items
Axis	Horizontal and vertical axes	A single-position encodes information by the placement of a mark on an axis.	Easy – Item 1 Moderate – Items 2, 3 Difficult – Items 4, 5
Apposed-position	Line chart, bar chart, plot chart	Information is encoded by a marked set that is positioned between two axes.	Easy – Item 6 Moderate – Item 7 Difficult – Item 8
Retinal-list	Graphics featuring colour, shape, size, texture, orientation	Retinal properties are used to encode information. These marks are not dependent on position.	Easy – Item 9 Moderate – Items 10, 11 Difficult – Item 12
Map	Road map, topographic map	Information is encoded through the spatial location of the marks.	Easy – Item 13 Moderate – Items 14, 15 Difficult – Items 16, 17
Connection	Tree, acyclic graph, network	Information is encoded by a set of node objects with a set of link objects.	Easy – Item 18 Moderate – Item 19 Difficult – Item 20
Miscellaneous	Pie chart, Venn diagram	Information is encoded with additional graphical techniques (e.g., angle, containment).	Easy – Item 21 Moderate – Item 22 Difficult – Item 23

Table 2
Means (and SD) for the six subtests (graphical languages) of the GLIM test by category of item difficulty

	Category					
Subtest	Easy		Moderate		Difficult	
Axis language	0.88 (0.33)	0.79 (0.41)	0.72 (0.43)	0.68 (0.47)	0.37 (0.48)	0.31 (0.46)
Apposed-position language	0.80 (0.40)	0.64 (0.48)	0.59 (0.49)	0.56 (0.50)	0.40 (0.49)	0.27 (0.45)
Retinal-list language	0.66 (0.47)	0.65 (0.48)	0.58 (0.49)	0.53 (0.50)	0.34 (0.48)	0.27 (0.45)
Map language	0.93 (0.26)	0.79 (0.41)	0.74 (0.42)	0.71 (0.45)	0.63 (0.48)	0.39 (0.49)
Connection language	0.91 (0.29)	0.76 (0.43)	0.61 (0.49)	0.48 (0.50)	0.25 (0.43)	0.21 (0.41)
Miscellaneous languages	0.96 (0.20)	0.76 (0.43)	0.70 (0.46)	0.68 (0.47)	0.60 (0.49)	0.46 (0.50)

Table 3
Means (and SD) for the six subtests (graphical languages) of the GLIM test by gender along with F values and effect sizes

	Total scores (Grades 4-6)				
Subtests	Male	Female	$F(1, 316)$	P	Cohen's d
Axis language	13.74 (2.79)	11.69 (3.06)	36.90	$p < .001$	.68
Apposed-position language	11.57 (3.01)	11.20 (2.83)	0.029	Ns	
Retinal-list language	11.21 (2.80)	10.50 (3.15)	1.92	Ns	
Map language	14.52 (2.30)	13.52 (2.56)	9.97	$p < .01$	.41
Connection language	11.12 (3.22)	10.74 (2.75)	0.153	Ns	
Miscellaneous languages	13.92 (2.80)	13.67 (3.19)	1.62	Ns	

Table 4
Means (and SD) for the six subtests (graphical languages) of the GLIM test by year and gender along with F values and effect sizes

	Year												
	Grade 4			Grade 5			Grade 6			Year		Gender	
Subtests	Male	Female	Total	Male	Female	Total	Male	Female	Total	F(2, 1032)	Partial η^2	F(1, 1032)	Partial η^2
Axis language	4.15 (1.17)	3.38 (1.34)	3.79 (1.31)	4.61 (1.24)	3.85 (1.44)	4.26 (1.38)	4.95 (1.33)	4.42 (1.17)	4.71 (1.78)	52.3***	.093	72.1***	.066
Apposed-position language	3.26 (1.40)	3.35 (1.28)	3.30 (1.35)	4.05 (1.32)	3.80 (1.20)	3.94 (1.27)	4.21 (1.26)	4.09 (1.24)	4.16 (1.25)	46.5***	.083	.04	.000
Retinal-list language	3.16 (1.36)	2.95 (1.35)	3.06 (1.36)	3.87 (1.24)	3.59 (1.40)	3.74 (1.32)	4.15 (1.26)	3.99 (1.42)	4.07 (1.39)	57.6***	.101	3.54	.003
Map language	4.34 (1.24)	4.09 (1.29)	4.23 (1.27)	5.08 (.94)	4.64 (1.09)	4.88 (1.03)	5.25 (.93)	4.99 (1.21)	5.13 (0.99)	71.5***	.122	16.48***	.016
Connection language	3.19 (1.39)	3.23 (1.19)	3.21 (1.30)	3.66 (1.35)	3.66 (1.17)	3.66 (1.27)	4.20 (1.27)	3.86 (1.22)	4.04 (1.26)	41.2***	.074	.11	.000
Miscellaneous languages	4.16 (1.41)	4.21 (1.45)	4.18 (1.43)	4.61 (1.29)	4.49 (1.25)	4.56 (1.27)	5.09 (.94)	4.97 (1.18)	5.03 (1.06)	47.2***	.084	.13	.000

*** $p < .001$.

Table 5
Means (and SD) for the three categories of item difficulty of the GLIM items by year and gender along with F values and effect sizes

	Year										
	Grade 4			Grade 5			Grade 6				
Categories	Male	Female	Total	Male	Female	Total	Male	Female	Total	F(2, 1032)	Cohen's d
Easy	9.39 (2.07)	9.20 (1.91)	9.30 (2.00)	10.16 (1.83)	9.99 (1.89)	10.08 (1.86)	10.75 (1.40)	10.66 (1.43)	10.71 (1.41)	.07	
Moderate	7.49 (2.59)	6.98 (2.39)	7.49 (2.59)	9.08 (2.19)	8.32 (2.35)	8.73 (2.29)	9.86 (2.12)	9.19 (2.39)	9.55 (2.28)	14.78 ***	.25
Difficult	5.37 (2.06)	5.02 (1.97)	5.37 (2.06)	6.64 (2.34)	5.74 (2.20)	6.23 (2.32)	7.22 (2.25)	6.47 (2.25)	6.87 (2.28)	19.00 ***	.28

*** $p < .001$.

Table 6
Means (and SD) of the GLIM moderate and difficult items which revealed statically significant gender differences

Item No. (see Appendix)	Graphical language	M (SD) Male	Female	F (1, 1050)	Cohen's d
7	Apposed-position	.67 (.47)	.59 (.49)	7.51***	.17
15	Map	.88 (.32)	.82 (.38)	7.56***	.17
2	Axis	.61 (.49)	.45 (.50)	27.78***	.32
11	Retinal-list	.67 (.47)	.52 (.50)	26.67***	.31
4	Axis	.48 (.50)	.29 (.45)	41.94***	.39
16	Map	.56 (.50)	.46 (.50)	9.97***	.20
5	Axis	.81 (.39)	.69 (.46)	22.71***	.29
17	Map	.77 (.42)	.63 (.48)	25.56***	.31
20	Connection	.26 (.44)	.19 (.39)	7.07***	.17

Note: *** b < .004